

# Bayesian Feed Forward Neural Network-Based Efficient Anomaly Detection from Surveillance Videos

M. Murugesan\* and S. Thilagamani

Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Tamil Nadu, India

\*Corresponding Author: M. Murugesan. Email: murugesonmkc0004@outlook.com

Received: 25 October 2021; Accepted: 02 January 2022

**Abstract:** Automatic anomaly activity detection is difficult in video surveillance applications due to variations in size, type, shape, and objects' location. The traditional anomaly detection and classification methods may affect the overall segmentation accuracy. It requires the working groups to judge their constant attention if the captured activities are anomalous or suspicious. Therefore, this defect creates the need to automate this process with high accuracy. In addition to being extraordinary or questionable, the display does not contain the necessary recording frame and activity standard to help the quick judgment of the parts' specialized action. Therefore, to reduce the wastage of time and labor, this work utilizes a Bayesian Feed-Forward Neural Network (BFFN) algorithm to automate the anomaly Recognition System. It aims to detect irregularities, regular signs of aggression, and violence automatically, rather than the conventional methods. This work uses Bayesian Feed Forward Neural Networks to identify and classify frames with high activity levels. From there, the proposed system can issue detection alerts for threats indicating suspicious activity at a particular moment. The simulation results show that the proposed BFFN based anomaly detection responds better than existing methods. The sensitivity, specificity, accuracy, and F1-score of the proposed system are 98.69%, 97.01%, 99.65%, and 0.98%.

**Keywords:** Anomaly; bayesian; feed-forward neural network; background modeling; clustering

## 1 Introduction

Human functional recognition can use various scenarios and detecting anomalies in security systems is one of them. An available recognition system is expected to identify the basic daily activities performed by a human being. It is trying to accomplish high accuracy for perceiving these activities based on the complexity and variety in human activities. Activity models needed to recognize and arrange human activities are developed depending on various methodologies explicit to the application. This research work discipline draws in video processing and machine learning classification. It discovers applications in a few studies like medication and health care, human-PC collaboration, crime investigation, and security frameworks. Utilizations of human movement recognition are not restricted to health care and security. It is a



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

continuous and open research point in PC vision that incorporates social biometrics, video investigation, animation, and union. Human activities, including signals and movements of the human body, are deciphered with the assistance of sensors. These understandings are fundamental for the recognizable proof of human activities.

The proposed system understands human activity by identifying movement and recognizing that move action. Initially, a pre-activity model is used to detect human activity accurately. This builds a complex conceptual model that can recognize and categorize all human activities. It uses a limited amount of captured sensor data to detect unknown patterns and does not require pre-samples for function mode detection. Although the two techniques contradict each other, they may have a consistent goal of achieving more excellent recognition performance for human functional recognition systems. These techniques complement each other, and they combine to improve performance using action method innovation that defines the recognition process.

## 2 Research Background

The most challenging feature of computer vision problems is synchronization to a specific event that is often detected by disorder over isolating a particular part. Only an image as a videotape is made up of spatial information, both spatial and kinetic data. Video analysis using both perceptions is available in considerable detail to recognize a particular activity, yet the extra information cycle requires huge assets to remove. It is necessary to plan an algorithm to delete information from the address of sensible period videos successfully. Several types of research have attempted to diagnose problems in films using carefully chosen highlights. The Edge Oriented Histogram (EOH) and Multi-layer Histogram of Optical Flow (MHOF) approach [1–3] are two suggested methods for detecting flaws. Another way makes advantage of the global example is by calculating Markovian differences from a neighboring example, with the time scale being displayed globally [4]. The Gaussian Mixture Model (GMM) process is used to authenticate cars. Then an average measure of computing location, speed, and bearing is used to predict whether a collision will occur [5,6]. Apart from recognizing anomalies in a specific crime, the general method of detecting anomalies is to use a spatiotemporal features identifier to distinguish the local descriptor of a particular event and train a classifier to get the global representation of the usual affair. Despite the high cost of computing, iDT is often regarded as the most acceptable approach among all hand-designed methods [7–10]. When used with a standard video dataset, these algorithms will identify low probability designs as anomalies.

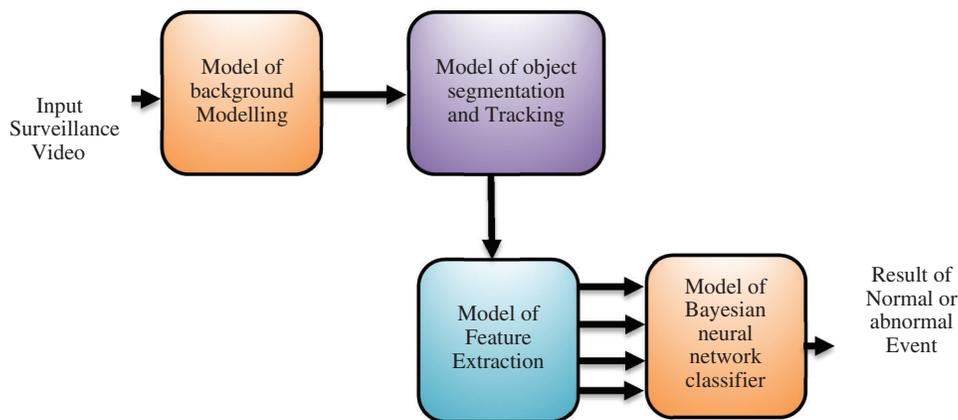
Due to the quick advancement of the processing framework, an in-depth learning approach was designed to classify an action into videos to learn characteristics. The basic deep learning approach for image processing is the convolutional neural network (CNN). The CNN Differences was built to deal with image processing difficulties in the video. A 3D CNN extracts external and period measurements using a 3D kernel [11,12]. The UCF-crime database is classified using this way [13,14]. Long Short-Term Memory is another way for obtaining unstable information that involves repeated formatting (LSTM). Another study [15,16] employs CNN with distinct spatial characteristics based on LSTM and a variation that integrates the findings into violent situations in movies. This study distinguishes all resilient practices based on anomalous phenomena. Differentiate between state-of-the-art video outcomes achieved using a two-stream-based structure [17,18]. The two data streams include spatial and temporal information [19–21]. After that, add the results together to produce a predicted score. For human anomaly detection, a recurrent neural network (RNN) is employed in conjunction with a Hidden Markov Model (HMM) in the combination of Fuzzy rule-based systems (FRBS) [22,23]. In [24], abnormalities relating to room occupancy are recognized using the Self-Organizing Map (SOM), whereas in [25], the Echo State Network (ESN) is used.

Researchers are learning the distribution of regular motions through training done using the maximum accessible recordings, according to the literature study thus far. They attempted to think of them as imperfections to recognize tiny structures. Even with sparse matrices, certain representations have shown to be quite successful in tackling challenges connected to computer scope. They're also thought to be the consequence of a significant reconfiguration mistake when tested, and specific patterns are inconsistent. As a result, in this paper, we provide a novel way to overcome the shortcomings of traditional methods.

### 3 Proposed Anomaly Detection System

This section explains how the proposed anomaly detection method and its many components operate. The video event categorization is done by examining the video stream's content. The technique's primary goal is to determine if the video/frame is abnormal or not.

The suggested system's block diagram is shown in Fig. 1. Background modeling, object segmentation, feature extraction, and classification are the four steps of the proposed anomaly detection system. The following stages describe the working procedure of the suggested anomaly detection system.



**Figure 1:** Proposed BFFN-based anomaly detection system block diagram

#### 3.1 Work Flow of Proposed Anomaly Detection

---

Input: Video Stream

Output: Classification of Normal or Abnormal Event

1: For each video clip

Obtain background:

Take pixel values in mode at each picture place. This is the setting. The pixel value that appears most often at a given picture is the mode.

Obtain foreground:

The foreground is obtained by subtracting the background from each image in the video clip.

Sharpen foreground image using a laplacian filter

Obtain black and white image:

---

(Continued)

---

(continued)

---

Take the absolute value of the foreground image. Convert to black and white using Matlab inbuilt function, which obtains the black & white photograph by thresholding.

Remove tiny regions that have less than 10 pixels to reduce noise.

Track objects: To track each object in the image, search in subsequent images with the video clip's given object.

vii. Extract feature: The features are extracted for each tracked object. The extracted features are as follows :

Subsequent frames are the distance between the midpoint in objects in x and y directions

Box size of material boundary in x and y directions

Several (white) pixels in the object.

Gray object image histogram limit the value of Background pixels; it can be divided into a total number of pixels in the object frame.

viii. Objects that exist in less than 10 frames are discarded

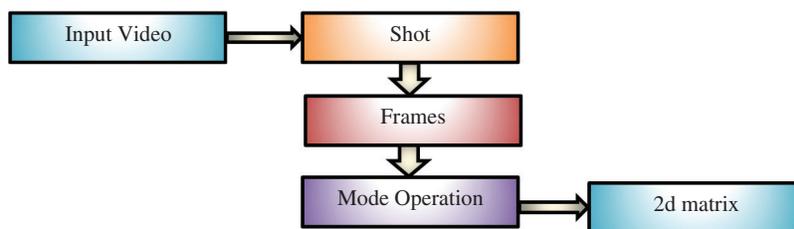
2: Classification Module :

After extracting features for all clips, Training and testing were done using a feed-forward neural network with trainbr as a training function. The Bayesian networks are incorporated to detect abnormality or Normal video events.

---

### 3.2 Background Modeling

The backdrop from the video frames is calculated using the model approach in this module. The video footage is first divided into frames, with the fundamental structures being identified using the shot segmentation approach. Wavelet Shot segmentation based on transforms and distance measures is used in this case. Following that, the backdrop is formed using the pixel value series mode at each picture position to extract the background. A functional block diagram of background modeling is shown in Fig. 2.



**Figure 2:** Block diagram of background modeling module

For example, video processing applications, summarization, synopsis, recovery, and video shot division are crucial tasks. The first info video sequence is initially divided into non-overlapping units considered shots that portray various actions. Each shot is characterized by no considerable changes in its substance controlled by the background and the items present in the scene.

Several techniques are available for video shot segmentation. In this work, Wavelet Transform and distance measure is used to distinguish the number of frames engaged with each shot. The first and

second frame is separated into blocks from the outset, and Discrete Wavelet Transform (DWT) is applied to each obstruct of the frame. Wavelets are functions that have an average value of zero in a defined restricted interval. Signifying any given function as a superposition of a set of wavelets or basis functions is the fundamental concept of wavelet transforms. The DWT obtains k sequences f x real numbers by transforming then sequences of m real numbers y1...ym allows,

$$Wt_{kx} = \int_{-\infty}^{+\infty} y(t)\psi_{kx}(t)dt \tag{1}$$

where

$$\psi_{kx}(t) = \frac{1}{\sqrt{C_{jk}}} \psi\left(\frac{t - d_{kx}}{a_{kx}}\right) \tag{2}$$

where  $\psi$ , c, and d are the mother wavelet, scaling parameter and shifting parameter. Then, the distance measure (dsij) is computed in-between frame 1 (fri) and frame 2(frj) using the following distance formula:

$$dS_{ij} = \sqrt{\sum_{j=1}^m (fr_i - fr_j)^2} \tag{3}$$

After finding the distance measure for the video’s first two frames, the procedure steps are used for the video’s consecutive frames. The frames within a shot can be identified by minimizing the distance measure, which compares two video frames. Let the original input video sequence be represented by Oiv [ j, k] and after shot segmentation, let the non-overlapping shots be represented by Nos[ j, k]. The function of background modeling comprises two stages, such as Background frame initialization and background subtraction.

For video analysis, the background frame initialization is one of the most important for starting background modeling. At the point when the background image is instated, it will be introduced as the reference image. There are numerous strategies to acquire the underlying background frame. For instance, compute the background image by averaging the main images, or accept the principal frame as the background straightforwardly or utilize background image successions without the possibility of moving items to appraise the background model boundaries. In this part, consider that the video starts with the background without moving objects. This work uses the selective averaging method to obtain the initial background model as:

$$B(s) = \frac{\sum_{m=1}^N I_m(S)}{N} \tag{4}$$

where

B(s) = background model’s Intensity

Im(S) = mth frame’s intensity.

N = Total number of frames

### 3.2.1 Background Subtraction

After obtaining the underlying background model, the deduction between the current frame and the reference frame is accomplished for the moving object distinguished. The deduction will be done pixel by pixel of the two frames. The basic form of this plan, where a pixel at an area in the current frame, is shown as foreground if:

$$|f_s - B_s| > \delta \quad (5)$$

where

BS = Background Image

$\delta$  = Adaptive threshold calculated by DWT

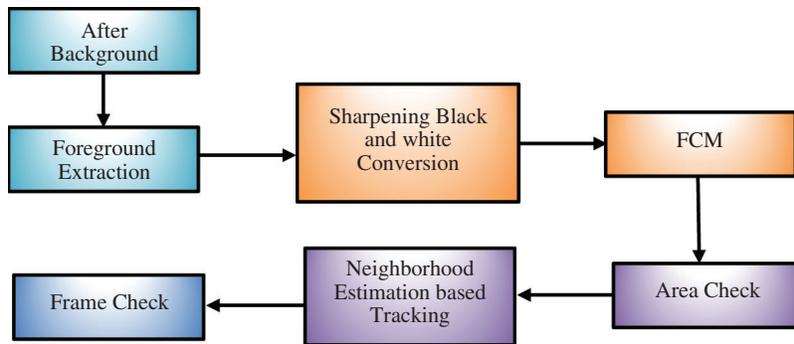
Background subtraction is obtained by distinguishing between the current image and the background image, which is also used to know the pixel intensity of the foreground. Diffs can be thought of as an image binary foreground.

$$Diff_s(k) = \begin{cases} 1 & \text{for } |f_s(k) - B_{(s)(k-\gamma)}| > \delta \\ 0 & \text{for others} \end{cases} \quad (6)$$

where  $\gamma$  represents the time interval of current and old frames.

### 3.3 Object Segmentation Using Fuzzy C-Means Clustering

The object is segmented using the Fuzzy C-Means operator in this module. The foreground is first retrieved, then sharpened and transformed to black and white. Following that, a Fuzzy C-means clustering process is performed, with regions smaller than a threshold being deleted. The picture is tracked by looking for each object's image to succeed from each video sequence using neighbourhood estimate among frames. Fig. 3 shows the object segmentation and tracking module's block diagram.



**Figure 3:** Object segmentation and tracking block diagram

Generally, the areas of interest in an image are objects (humans, cars, text, etc.) in its foreground. The motivation of the “background pattern” proposed path to detect moving objects in the current frame and frame is usually distinguished from the “background image”. In the proposed case, frame differencing is used as the foreground detection technique. Let the input video sequence be defined as  $Vid[x, y, t]$ , where  $x$  and  $y$  define the pixel position and  $t$  gives the time. Then the frame difference at a time ‘ $t+1$ ’ is given by.

$$F_{dis}(t+1) = |Vid[x, y, t+1] - Vid[x, y, t]| \quad (7)$$

The background is to be the frame at time ‘ $t$ ’. After obtaining the foreground, it is sharpened by the use of a Laplacian filter. The discrete Laplacian is characterized as “the amount of the second subsidiaries Laplace administrator organize articulations and determined as the number of contrasts over the closest neighbors of the central pixel”. Like this, it is changed over to the grayscale design. A grayscale

computerized picture is “a picture where the estimation of every pixel is a solitary example, that is, it conveys just intensity information”. Images of this kind, often known as black-and-white, range in intensity from dark at the lowest to white at the highest. In segmentation, there are various clustering techniques, and Fuzzy C-means (FCM) is one of them. When compared to the K-means clustering approach, FCM produces more accurate clustering results using fuzzy notions. The objective minimization function of FCM is defined by:

$$S_i = \sum_{i=1}^N \sum_{j=1}^{cen} m_{ij}^n |Z_i - cen_j|^2 \quad (8)$$

Here,  $\mu_{ij}$  is the degree of membership function of  $i$ th Cluster  $j$  has  $D$ -dimensional measured data ( $D_i$ ). The centroid of  $j$ th cluster is represented by  $cen_j$ . On every iteration, By modifying the membership function, the FCM's minimization function is optimized for clustering  $\mu_{ij}$  and centroid  $cen_j$ . The updating equations are given by:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{cen} \left( \frac{\|z_i - cen_j\|}{z_i - cen_k} \right)^{\frac{2}{m-1}}} \quad (9)$$

where ‘ $z_i$ ’ is the input data, ‘ $cen$ ’ is the centroid and ‘ $m$ ’ is a positive constant. The examination of the centroid is  $cen_j$ , and other centroids are represented by  $cen_k$ . The centroid updating equation is given by:

$$cen_j = \frac{\sum_{i=1}^N \mu_{ij} Z_i}{\sum_{i=1}^N \mu_{ij}} \quad (10)$$

The iterative optimization procedure is carried out till it satisfies the equation

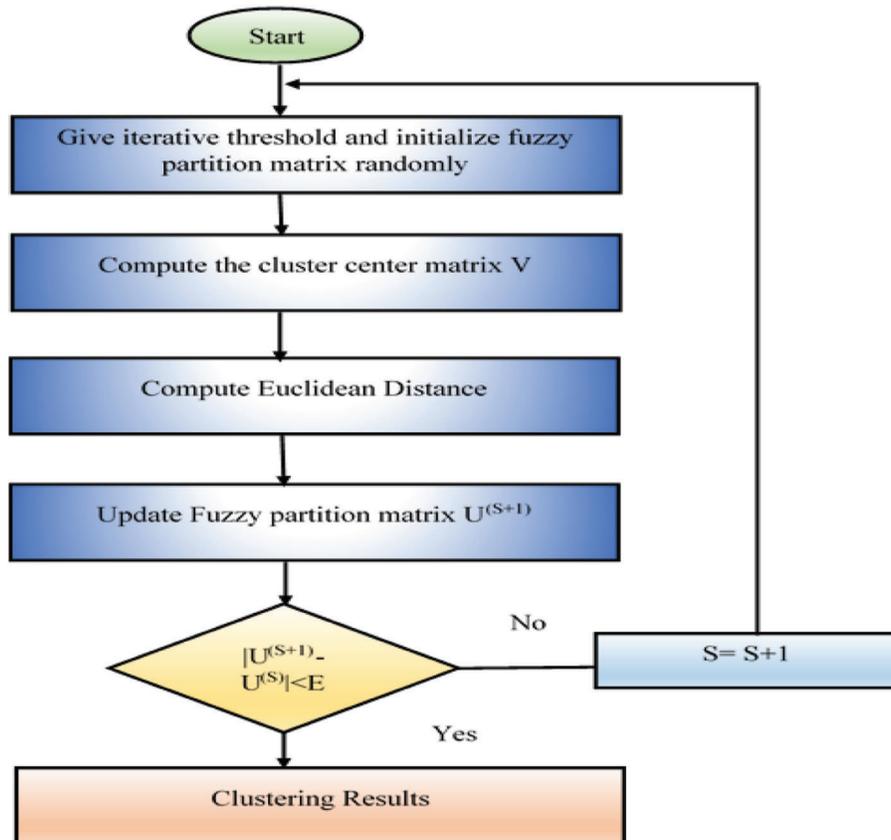
$$\text{maximum}_{ij} \left\{ \left| \mu_{ij}^{(k+1)} - \mu_{ij}^{(k)} \right| \right\} < \varepsilon \quad (11)$$

where  $\varepsilon$  is the termination condition whose value is between 0 and 1. The iteration is continued till the local minima or saddle point for the objective function is found. Many segments are formed with FCM results, and some of the segments are irrelevant. To remove these sections, those parts that have less area are considered. For this, a threshold is fixed, and the region in each segment is analyzed. The part territory is the complete number of pixels in the relating segment. Assume the limit set is addressed by the space of Thr and the  $i$ th segment, and then the condition is defined as:

$$\text{if } (Ari < Thr), \text{ then remove Ari} \quad (12)$$

$$\text{if } (Ari > Thr), \text{ then accept Ari} \quad (13)$$

Exclusion of smaller areas also falls out in reducing noise. Using neighborhood estimation among frames, each object is tracked by searching subsequent images with the video clip's given object. Subsequently, the objects that exist in less than 10 frames are discarded. Fig. 4 shows the flowchart of fuzzy C-means clustering. The FCM algorithm will give better quality than the K-means algorithm, where the k-means data focus should exclusively have one cluster center. As a result of the data point being assigned to each cluster community, the information point may be associated with more than one cluster centre.



**Figure 4:** Flow chart of fuzzy C-means clustering

### 3.4 Feature Extraction

From the tracked objects, this module outputs characteristics such as distance measure, size, number of pixels, and histogram. The Distance measure feature is determined by the distance in x and y dimensions between the centroid of objects in consecutive frames of the video clip. The centroids of the objects are first determined in each frame. Let  $Ob$  represents the objects which are present in the frame. Let the pixels inside the  $i$ th object be represented by  $poi_j$ , where  $0 < j < n$  and the corresponding centroid are calculated as follows:

$$cen_i = \frac{\sum_{j=1}^n poi_j}{n} \quad (14)$$

Subsequent frames are presented in the midpoint of the material after each object is found midpoint in the Euclidean distance from the frame. Thus the value of the distance “ is measured by the distance to the feature distance. This module’s subsequent component is an item’s size with a bounding box in x and y headings. For anomaly situations, the bounding box for two measurements (x and y) is discovered. It means “measure every available box in the pixels inside it with dots”. It is additionally referred to as the “base perimeter bounding box”. The third component taken is the number of pixels. The shape territory of the object serves as an element of the number of pixels enclosed within the limit of the item. The fourth and last element considered is the histogram self-esteem. The histogram is “a graphical depiction of the allocation of pixels of a picture”. The histogram addresses “the frequency of the tissue, appearing as a nearby square shape, raised in discrete segments, and the space corresponding to the perceived frequency in the time interval”. The histogram peak is the most extreme histogram self-esteem in the shooting span. The histogram self-esteem is the total number of pixels separated by “histogram peaks in the target image that eliminate the wake-up of background pixels.

### 3.5 Anomaly Classification

The frames are classified as normal or abnormal in this module using a feed-forward Bayesian Neural Network. The video clip library is split into two parts: 60 percent for training and 40 percent for testing in neural networks. Training and testing are carried out when all of the early modules of backdrop modelling, object segmentation, and object tracking have been completed.

The Artificial Neural Networks Classifier provides a powerful tool for analyzing, modeling and predicting. In post-theoretical facts, the benefits of the nervous system are presented. First, “neural networks are self-adaptive methods driven by the data themselves, with no explicit description of the unique model dynamic or spreadable form they can fine-tune”. Second, there are “universal functional estimates in which the neural network can approximate any function with random accuracy”. There are “non-linear models”, which makes the nervous system enlarge in forming complex real-world relationships. The posterior probabilities, which form the basis of neural network classification rules and statistical analysis, can be estimated. The nervous system consists of three layers: the input layer, the hidden layer, and the output layer. The nervous system works in two stages: the training phase and the other testing stage

Traditionally, the backpropagation algorithm is followed by neural networks. There has been a lot of research on ways to speed up the convergence of the algorithm. This learning rate has been used for changing momentum and re-conducting variables. The overweight system applications focus on sequential deployments where the input/output pairs are updated after each presentation. The feed-forward network is becoming more popular as it closes the flaws of the back transmission algorithm. The three-layer feed-forward network is given in Fig. 5.

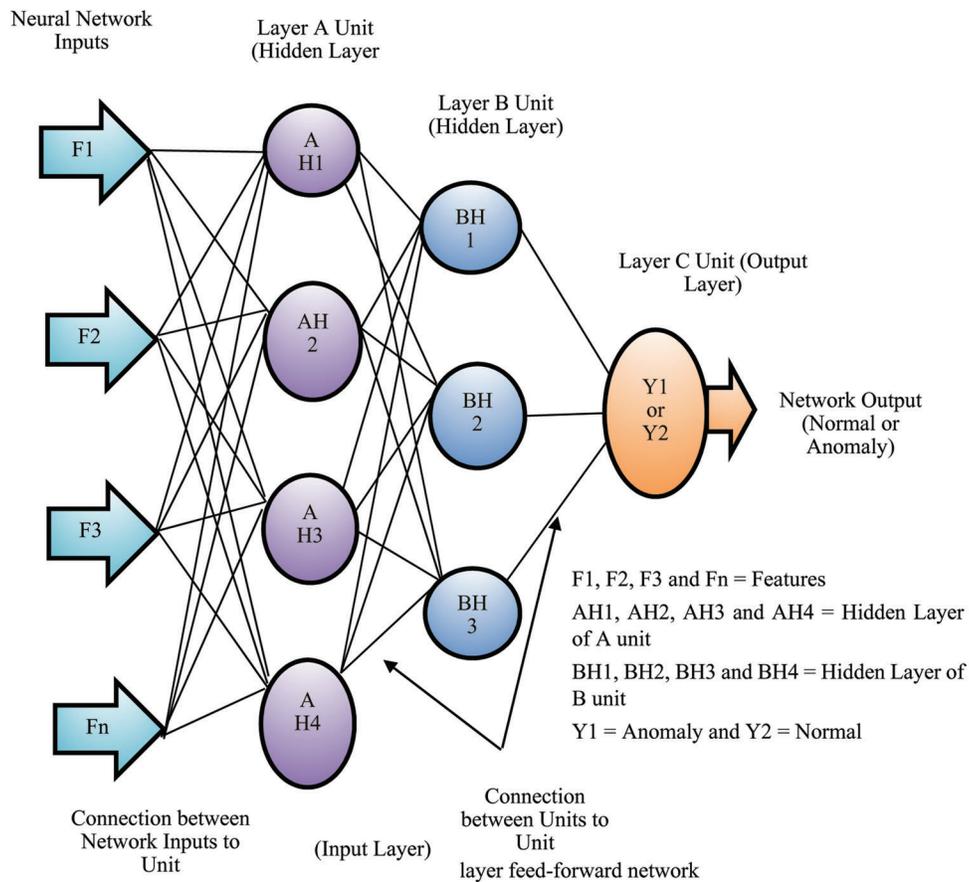


Figure 5: A three-layer feed-forward network

---

 Algorithm of Bayesian feed-forward neural network
 

---

Step 1: Let the net input to  $i$ th unit in  $(m+1)$ th layer be represented as:

$$inp^{k+1}(m) = \sum_{i=1}^{sk} \bar{\omega}^{m+1}(m, i) y^k(i) + x^{k+1}(m) \quad (15)$$

Step 2: The output of the  $m$ th unit can be defined by:

$$x^{k+1}(m) = G^{k+1}(inp^{k+1}(m)) \quad (16)$$

Step 3: The main function of the network is to generate input and to know the set relationship of pairs  $\{(g_1, \tau_1), (g_2, \tau_2), \dots, (g_p, \tau_p)\}$ .

Step 4: The performance index for the network is given by:

$$F = \frac{1}{2} \sum_{p=1}^P (\tau_p - y_p^{N_s})^T (\tau_p - y_p^{N_s}) = \frac{1}{2} \sum_{p=1}^P er_p^T er_p \quad (17)$$

Step 5: Here,  $y_p^{N_s}$  is the output for  $p$ th input and  $er_p = \tau_p - y_p^{N_s}$  is the error for  $p$ th input. In backpropagation, The performance index can be interpreted as approximated by the steep descent rule:

$$F = \frac{1}{2} er_p^T er_p \quad (18)$$

Step 6: In the steepest descent approximation, the squares' total is the same input-output as the pair of squared errors. Hence,

$$\Delta \bar{\omega}^k(m, i) = -\alpha \frac{\partial F}{\partial \bar{\omega}^k(m, i)} = -\alpha \gamma^k(m) \gamma^{k-1}(i) \quad (19)$$

$$\Delta \gamma^k(m) = -\alpha \frac{\partial F}{\partial Z^k(m)} = -\alpha \gamma^k(m) \quad (20)$$

Where,

$$\gamma^k(m) = \frac{\partial F}{\partial inp^k(m)} \quad (21)$$

Step 7: Here,  $\alpha$  is the learning rate, and  $\gamma$  is the sensitivity of the performance index. It can also show up in terms of sensitivity

$$\gamma^k = G^k(inp^k) \bar{\omega}^{k+1} \gamma^{k+1} \quad (22)$$


---

The Bayesian network represents the dependency among random variables and is a possible logical model that provides a brief description of the joint probability distribution. The Bayesian networks are incorporated into feed-forward networks to have the Bayesian-based feed-forward neural network.

The general learning algorithm presently continues as follows; first, increase the contribution forward utilizing Eqs. (16) and (17); next, propagate the sensitivities back utilizing Eq. (19) lastly, update the loads and balances utilizing Eqs. (21) and (22). Approximately 80% of the frames are utilised for learning, while the remaining 20% are used for assessment. The first phase of the three modules is used to pre-process both learning and testing networks. Following the learning step, the network is given test data frames to categorise them as normal or anomalous.

#### 4 Simulation Results and Discussion

The simulation results and performance analysis of the suggested anomaly detection system are discussed in this section. UCSD anomaly detection datasets are utilised in the suggested approach of experimentation. A fixed camera situated at a height above pathways was used to create this dataset. The group thickness in the pathways varied from thick to highly jammed. In a normal context, however, the video just shows in Fig. 6 walkers. The flow of non-walker substances in the pathways or unconventional person-on-foot movement designs are both rare occurrences. ‘Bikers,”skaters,”small trucks,’ and ‘individuals wandering over a path or in the grass around it’ are all common occurrences. Matrix laboratory (MATLAB) software is used to accomplish the suggested method.

**Frames 1 to 4**



**Frames 5 to 8**



**Frames 9 to 12**



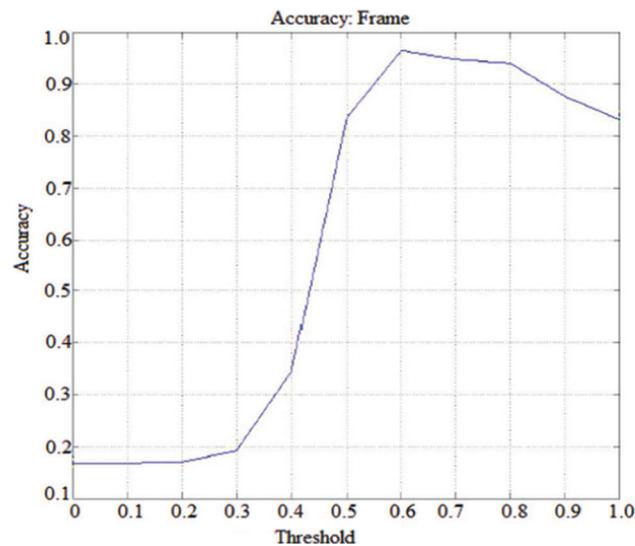
**Frames 13 to 16**



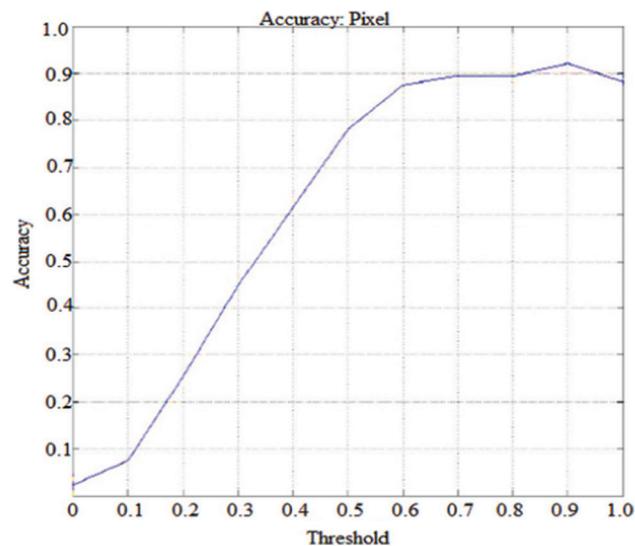
**Figure 6: Simulation results**

This section talks about the suggested anomaly detection system’s simulation findings and performance analyses. UCSD anomaly detection datasets are employed in the suggested approach of experimentation. A fixed camera situated at a height above pathways was used to create this data collection. The group thickness varied throughout the pathways, ranging from thick to highly clogged. The film, however, features just walkers in the regular scenario. The flow of non-walker substances in the pathways or odd person-on-foot movement designs are rare occurrences. ‘Bikers,”skaters,”small trucks,’ and ‘individuals wandering over a path or in the grass around it’ are all examples of often occurring oddities. Matrix laboratory (MATLAB) software is used to accomplish the suggested methodology.

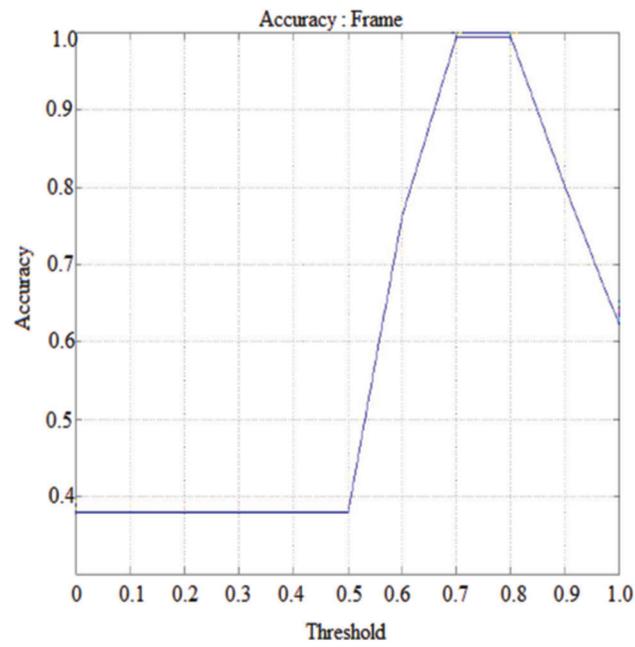
Figs. 7–10 gives the curves obtained for the proposed technique. Figs. 7 and 8 give the accuracy curve of Peds1. Figs. 8 and 9 give the accuracy curve of Peds2. The accuracy graph is plotted for both the frame case and pixel case for both data sets. In Peds1, the maximum accuracy obtained is 0.92 for pixel (at threshold 0.9) and 0.96 for the frame (at threshold 0.6). In Peds2, the maximum accuracy obtained is 0.99 for the frame (at threshold 0.8) and 0.85 for pixel (at threshold 0.8). The ROC curves for the proposed approach and the other technique are shown in Fig. 11. The blue line in the ROC graph represents the suggested technique's average ROC curve, whereas the red line represents the alternative technique's average ROC curve. Fig. 10 shows that the suggested strategy has a superior ROC curve than the previous method.



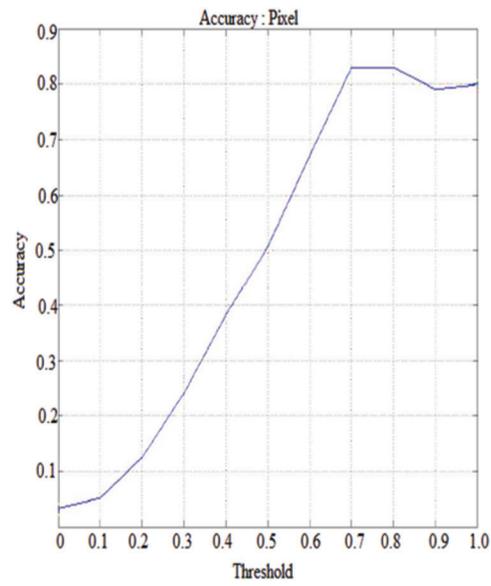
**Figure 7:** Accuracy (of frame) for Peds1



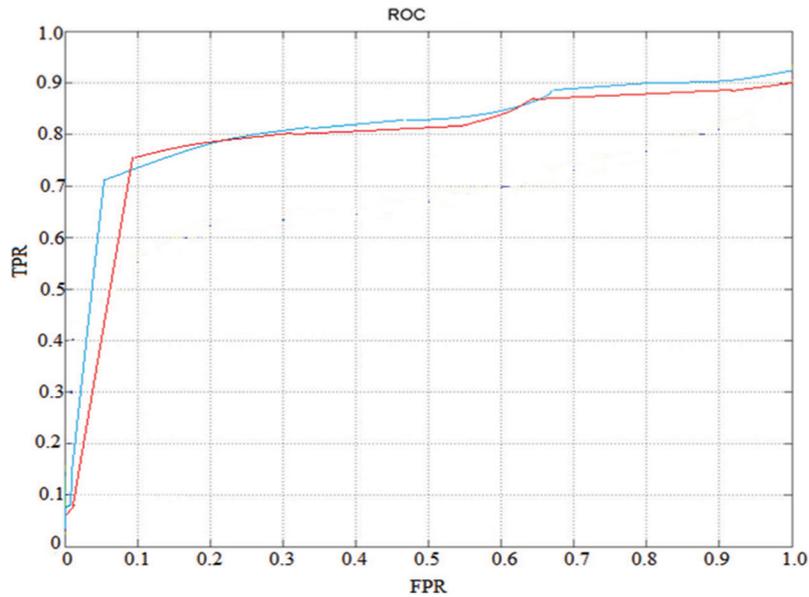
**Figure 8:** Accuracy (of a pixel) for Peds1



**Figure 9:** Accuracy (of frame) for Ped2

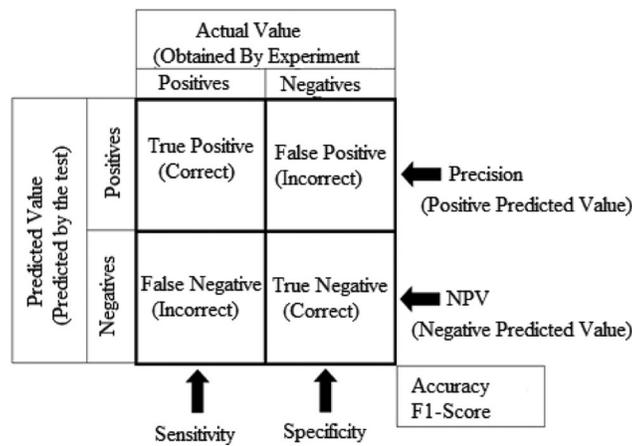


**Figure 10:** Accuracy (of a pixel) for Ped2



**Figure 11:** ROC curves

The performance analysis of the proposed system is validated using different metrics, such as sensitivity, specificity, accuracy and F1-score. In these parameters obtained from the confusion matrix, the diagrammatic representation of the confusion matrix is shown in Fig. 12. The mathematical formulation of sensitivity, specificity, accuracy and F1-score are represented by Eqs. (20)–(23).



**Figure 12:** Model of confusion matrix

Sensitivity: Test sensitivity is its ability to determine cases of anomaly accurately. To evaluate it, we need to calculate the true positive ratio in cases of anomaly. Mathematically, this can be specified

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \tag{23}$$

Specificity: The specificity of a test is characterized differently, regularly. For example, particularity is the capacity of a screening test to recognize a true negative depends on the true negative rate.

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \tag{24}$$

Accuracy: Accuracy is the approximate value of the algorithm class showing the actual value probability of the label (estimating the overall effectiveness of the algorithm).

$$\text{Accuracy} = \text{TN} + \text{TP} / (\text{TN} + \text{TP} + \text{FN} + \text{FP}) \tag{25}$$

F1-score: The F-score is a combination of high sensitivity covering protocols and high accuracy that will challenge those with higher specificity.

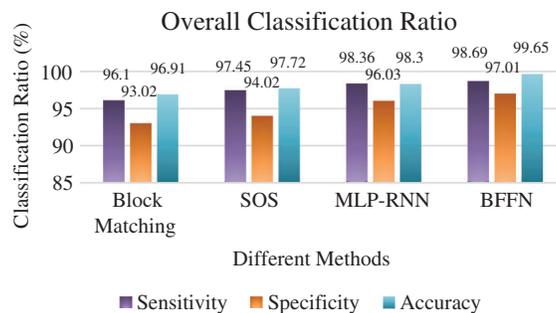
$$\text{F1 - Score} = \text{TP} / (\text{TP} + 1/2(\text{FP} + \text{FN})) \tag{26}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

Tab. 1 and Fig. 13 discuss the Overall Classification Ratio Performance analysis for proposed BFFN based anomaly detection with existing Block matching, SOS and MLP\_RNN methods. In this comparison, it is obvious that the proposed BFFN-based anomaly detection system performs well on all video sequences. The proposed system’s sensitivity, specificity, and accuracy are 98.69%, 97.01% and 99.65%.

**Table 1:** Comparative analysis of existing Covid -19 forecasting approaches

| Different anomaly detection methods                       | Anomaly detection sensitivity (%) | Anomaly detection specificity (%) | Anomaly detection accuracy (%) | F1-Score |
|---|-----------------------------------|-----------------------------------|--------------------------------|----------|
| Block matching  | 96.10                             | 93.02                             | 96.91                          | 0.89     |
| Sequential Outline Separation (SOS)                       | 97.45                             | 94.02                             | 97.72                          | 0.91     |
| Multi-Layer Perceptron Recurrent Neural Network (MLP_RNN) | 98.36                             | 96.03                             | 98.30                          | 0.93     |
| Proposed Bayesian Feed Forward Neural Network (BFFN)      | 98.69                             | 97.01                             | 99.65                          | 0.98     |



**Figure 13:** Performance analysis of overall classification ratio

## 5 Conclusion

In this research work, a new approach has been developed using content-based features for video anomaly detection in surveillance videos. A Bayesian feed-forward neural network is used for video

event classification in this approach. Background modelling, object segmentation and tracking, feature extraction, and classification are the four components that make up the proposed technique. The UCSD Anomaly Detection Datasets would be used in the implementation. Receiver Operating Characteristics (ROC) curves and classification accuracy are used to assess the suggested anomaly detection system's performance. The curves are taken for both frame case and pixel case for Peds1 and Peds2 of the UCSD dataset. The maximum accuracy came to 0.92 for pixel and 0.99 for the frame. The proposed technique also achieved good ROC curves for all conditions. The sensitivity, specificity, and accuracy of the proposed anomaly detection method are 98.69%, 97.01% and 99.65%. The proposed method is used only for anomaly detection in the video streams but, not suitable for human action classification. A deep learning algorithm-based anomaly detection approach is proposed in future work to recognize and classify the different types of anomaly actions to overcome the issues of the BFFN method.

**Funding Statement:** The authors received no specific funding for this study

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] W. Sultani, C. Chen and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 6479–6488, 2018.
- [2] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1436–1452, 2013.
- [3] X. Zhu, J. Liu, J. Wang, C. Li and H. Lu, "Sparse representation for robust abnormality detection in crowded scenes," *Pattern Recognition*, vol. 47, no. 5, pp. 1791–1799, 2014.
- [4] Y. Benezeth, P. M. Jodoin, V. Saligrama and C. Rosenberger, "Abnormal events detection based on spatio-temporal co-occurrences," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, pp. 2458–2465, 2009.
- [5] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, pp. 2921–2928, 2009.
- [6] W. Li, V. Mahadevan and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.
- [7] R. T. Ionescu, S. Smeureanu, B. Alexe and M. Popescu, "Unmasking the abnormal events in video," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 2914–2922, 2017.
- [8] A. Del Giorno, J. A. Bagnell and M. Hebert, "A discriminative framework for anomaly detection in large videos," in *Proc. of the European Conf. on Computer Vision*, Amsterdam, The Netherlands, pp. 334–349, 2016.
- [9] H. Gao, W. Huang and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intelligent Automation & Soft Computing*, vol. 25, no. 3, pp. 547–559, 2019.
- [10] W. Brendel, A. Fern and S. Todorovic, "Probabilistic event logic for interval-based event recognition," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, pp. 3329–3336, 2011.
- [11] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Providence, RI, USA, pp. 2112–2119, 2012.
- [12] M. Bertini, A. Del Bimbo and L. Seidenari, "Multi-scale and real-time non-parametric approach for anomaly detection and localization," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 320–329, 2012.

- [13] R. Leyva, V. Sanchez and C. T. Li, "Video anomaly detection with compact feature sets for online performance," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3463–3478, 2017.
- [14] Y. Cong, J. Yuan and Y. Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 10, pp. 1590–1599, 2013.
- [15] C. Lu, J. Shi and J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Sydney, Australia, pp. 2720–2727, 2013.
- [16] S. Biswas and R. V. Babu, "Real time anomaly detection in H.264 compressed videos," in *proc. of the national conf. on computer vision*, in *Pattern Recognition, Image Processing and Graphics*, Jodhpur, India, pp. 1–4, 2013.
- [17] A. Adam, E. Rivlin, I. Shimshoni and D. Reinitz, "Robust real-time unusual event detection using multiple fixed- location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.
- [18] F. Mallouli, "Robust em algorithm for iris segmentation based on mixture of Gaussian distribution," *Intelligent Automation & Soft Computing*, vol. 25, no. 2, pp. 243–248, 2019.
- [19] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, pp. 1446–1453, 2009.
- [20] K. W. Cheng, Y. T. Chen and W. H. Fang, "Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5288–5301, 2015.
- [21] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 733–742, 2016.
- [22] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in *Proc. of the Int. Symp. in Neural Networks*, Sapporo, Hakodate, and Muroran, Hokkaido, Japan, pp. 189–196, 2017.
- [23] K. Xu, T. Sun and X. Jiang, "Video anomaly detection and localization based on an adaptive intra-frame classification network," *IEEE Transactions on Multimedia*, vol. 22, no. 2, pp. 394–406, 2020.
- [24] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei *et al.*, "Spatial-temporal convolutional neural networks for anomaly detection and localization in crowded scenes," *Signal Processing: Image Communication*, vol. 47, pp. 358–368, 2016.
- [25] R. Mehran, A. Oyama and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Miami, Florida, USA, pp. 935–942, 2009.